

*INS
AI*

METHODS FOR USING MUTANT RNA POLYMERASES WITH REDUCED DISCRIMINATION BETWEEN NON-CANONICAL AND CANONICAL NUCLEOSIDE TRIPHOSPHATES

Field of Invention

The field of the present invention is methods for producing nucleic acid molecules containing at least one non-canonical nucleotide and for characterizing nucleic acid molecules by synthesizing nucleic acid molecules containing at least one non-canonical nucleotide *in vitro* using mutant nucleic acid polymerases having at least a 10-fold reduced discrimination between 2'-deoxyribonucleoside-5'-triphosphates and ribonucleoside-5'-triphosphates as substrates compared to the corresponding wild-type enzymes.

Background

There are a number of procedures commonly used in the art for *in vitro* synthesis of nucleic acid molecules, including both DNA and RNA. For example, one may use an *in vitro* transcription reaction to synthesize RNA from a DNA template present in the reaction. T7-type RNA polymerases, such as T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase, are commonly used in such reactions, although many other RNA polymerases may also be used. Usually, but not always, synthesis of RNA is *de novo* (i.e., unprimed), and usually, but not always, transcription is initiated at a sequence in the template that is specifically recognized by the RNA polymerase, termed a "promoter" or a "promoter sequence". A method for *in vitro* transcription is presented herein.

Procedures for *in vitro* nucleic acid synthesis are also commonly used in the art to amplify nucleic acid molecules, including both DNA and RNA. For example, transcriptions using RNA polymerases are an integral part of "nucleic acid

DRAFT EDITION

sequence-based amplification" (NASBA), "self-sustained sequence replication" (3SR), and "transcription-mediated amplification" (TMA) Hill, C.S., 1996, three similar methods for amplifying nucleic acid molecules *in vitro*.

5 By way of example, all or a specific portion of an RNA molecule may be amplified using NASBA (Compton, *et al.*, 1991) or 3SR (Fahy, *et al.*, 1991) by isothermal incubation of a sample RNA in a buffer containing two primers (a first primer complementary to the RNA molecule and encoding a promoter sequence for an RNA polymerase and a second primer complementary to the 3'-end of the first cDNA strand resulting from reverse transcription of the RNA molecule), an RNA- and DNA-dependent DNA polymerase which also has RNase H activity (or a separate RNase H enzyme), all four canonical 2'-deoxynucleoside-5'-triphosphates (dATP, dCTP, dGTP and dTTP), an RNA polymerase that recognizes the promoter sequence of the first primer, and all four canonical ribonucleoside-5'-triphosphates (rATP, rCTP, rGTP and rUTP).

20 A first cDNA strand is synthesized by extension of the first primer by reverse transcription. Then, the RNase H digests the RNA of the resulting DNA:RNA hybrid, and the second primer primes synthesis of the second cDNA strand. The RNA polymerase then transcribes the resultant double-stranded DNA (ds-DNA) molecule from the RNA polymerase promoter sequence, making many more copies of RNA, which in turn, are reversed transcribed into cDNA and the process begins all over again. This series of reactions, from ds-DNA through RNA intermediates to more ds-DNA, continues in a self-sustained way until reaction components are exhausted or the enzymes are inactivated. DNA samples can also be amplified by other variations of NASBA or 3SR or TMA.

Another nucleic acid amplification method involving DNA synthesis is the polymerase chain reaction (PCR).

By way of example, a specific portion of a DNA molecule may be amplified using PCR by temperature cycling of a sample DNA in a buffer containing two primers (one primer complementary to each of the DNA strands and which, together, flank the DNA sequence of interest), a thermostable DNA polymerase, and all four canonical 2'-deoxynucleoside-5'-triphosphates (dATP, dCTP, dGTP and dTTP). The specific nucleic acid sequence is geometrically amplified during each of about 30 cycles of denaturation (e.g., at 95°C), annealing of the two primers (e.g., at 55°C), and extension of the primers by the DNA polymerase (e.g., at 70°C), so that up to about a billion copies of the nucleic acid sequence are obtained. RNA may be similarly amplified using one of several protocols for (reverse transcription-PCR) RT-PCR, such as, for example, by carrying out the reaction using a thermostable DNA polymerase which also has reverse transcriptase activity (Myers and Gelfand, 1991).

The polymerase chain reaction, discussed above, is the subject of numerous publications and patents, including, for example: Mullis, K.B., et al., U.S. Pat. No. 4,683,202 and U.S. Pat. No. 4,965,188.

A variety of procedures for using *in vitro* nucleic acid synthesis for characterizing nucleic acid molecules, including both DNA and RNA, also are known in the art.

There are many reasons for characterizing nucleic acid molecules. For example, genes are rapidly being identified and characterized which are causative or related to many human, animal and plant diseases. Even within any particular gene, numerous mutations are being identified

00000000000000000000000000000000

that are responsible for particular pathological conditions. Thus, although many methods for detection of both known and unknown mutations have been developed (e.g., see Cotton, 1993), our growing knowledge of human and other genomes 5 makes it increasingly important to develop new, better, and faster methods for characterizing nucleic acids. Besides diagnostic uses, improved methods for rapidly characterizing nucleic acids will also be useful in many other areas, including human forensics, paternity testing, animal and 10 plant breeding, tissue typing, screening for smuggling of endangered species, and biological research.

One of the most informative ways to characterize a DNA molecule is to determine its nucleotide sequence. The most commonly used method for sequencing DNA at this time 15 (Sanger, et al., 1977) uses a DNA polymerase to produce differently sized fragments depending on the positions (sequence) of the four bases (A = Adenine; C = Cytidine; G = Guanine; and T = Thymine) within the DNA to be sequenced. In this method, the DNA to be sequenced is used as a 20 template for *in vitro* DNA synthesis. RNA may also be used as a template if a polymerase with RNA-directed DNA polymerase (i.e., reverse transcriptase) activity is used. In addition to all four of the deoxynucleotides (dATP, dCTP, dGTP and dTTP), a 2',3'-dideoxynucleotide is also included 25 in each *in vitro* DNA synthesis reaction at a concentration that will result in random substitution of a small percentage of a normal nucleotide by the corresponding dideoxynucleotide. Thus, each DNA synthesis reaction yields a mixture of DNA fragments of different lengths 30 corresponding to chain termination wherever the dideoxynucleotide was incorporated in place of the normal deoxynucleotide.

The DNA fragments are labelled, either radioactively or non-radioactively, by one of several methods known in the art and the label(s) may be incorporated into the DNA by extension of a labelled primer, or by incorporation of a 5 labelled deoxy- or dideoxy- nucleotide. By carrying out DNA synthesis reactions for each of the four dideoxynucleotides (ddATP, ddCTP, ddGTP or ddTTP), then separating the products of each reaction in adjacent lanes of a denaturing polyacrylamide gel or in the same lane of a gel if different 10 distinguishable labels are used for each reaction, and detecting those products by one of several methods, the sequence of the DNA template can be read directly.

Radioactively-labelled products may be read by analyzing an exposed X-ray film. Alternatively, other methods commonly 15 known in the art for detecting DNA molecules labelled with fluorescent, chemiluminescent or other non-radioactive moieties may be used.

Because 2',3'-dideoxynucleotides (ddNTPs), including even ddNTPs with modified nucleic acid bases, can be used as 20 substrates for many DNA polymerases, Sanger's dideoxy- sequencing method is widely used. Recently, Tabor and Richardson (EP application 942034331, 1994) reported that mutations at specific sites in many DNA polymerases improved the ability of these mutant enzymes to accept ddNTPs as 25 substrates, thereby leading to improved DNA polymerases for DNA sequencing using the Sanger method.

Nucleic acid sequencing provides the highest degree of certainty as to the identity of a particular nucleic acid. Also, nucleic acid sequencing permits one to detect 30 mutations in a gene even if the site of the mutation is unknown. Sequencing data may even provide enough information to permit an estimation of the clinical

DRAFT

significance of a particular mutation or of a variation in the sequence.

Cycle sequencing is a variation of Sanger sequencing that achieves a linear amplification of the sequencing signal by using a thermostable DNA polymerase and repeating chain terminating DNA synthesis during each of multiple rounds of denaturation of a template DNA (e.g., at 95°C), annealing of a single primer oligonucleotide (e.g., at 55°C), and extension of the primer (e.g., at 70°C).

Other methods for sequencing nucleic acids are also known besides the Sanger method. For example, Barnes described a method for sequencing DNA by partial ribo-substitution (Barnes, W.M., 1977). In this method, a pre-labelled primer was extended *in vitro* on a template DNA to be sequenced in each of four reactions containing a wild-type DNA polymerase in the presence of Mn²⁺, all four canonical 2'-deoxyribonucleoside triphosphates, and one of four ribonucleoside triphosphates under deoxy-/ribo-nucleotide ratios and conditions that result in about 2% ribonucleotide substitution at each position. After alkali treatment to cleave the synthetic DNA at the positions of partial ribosubstitution, the sequence was determined by analyzing the fragments resulting from each reaction following electrophoresis on a denaturing polyacrylamide gel.

Although most methods for sequencing nucleic acids employ DNA polymerases, some work has also been reported on the use of T7 RNAP and SP6 RNAP for transcription sequencing of DNA templates beginning at the respective T7 or SP6 promoter sequence using 3'-deoxyribonucleoside-5'-triphosphates (Axelrod, V.D., and Kramer, F.R., 1985), and Q-Beta replicase for sequencing single-stranded RNA

DRAFT - DRAFT - DRAFT - DRAFT -

templates (Kramer, F.R., and Mills, D.R., 1978). Also, 3'-O-methyl-ribonucleoside-5'-triphosphates have been used for sequencing DNA templates with *E. coli* RNA polymerase ((Axelrod, V.D., et al., 1978). None of these techniques is commonly used at present, perhaps in part, due to the difficulty to obtain the 3'-deoxy- and 3'-O-methyl-nucleoside triphosphate substrates, while 2',3'-dideoxy-ribonucleoside-5'-triphosphates that are commercially available have not been found to be substrates for wild-type (w.t.) RNA polymerases.

In view of the numerous applications involving *in vitro* nucleic acid synthesis known in the art, it is useful to consider the properties of the key nucleic acid polymerase reagents which make these procedures possible, and which, if modified in their essential properties, might improve these procedures.

One classification of nucleic acid polymerases relies on their different template specificities (RNA or DNA), substrate specificities (rNTPs or dNTPs), and mode of initiation (*de novo* or primed). These designations usually refer to the template and substrate specificities displayed *in vivo* during the fulfillment of a polymerase's biological function.

In vitro, polymerases can display novel activities, albeit with reduced efficiency and/or under non-physiological conditions. *E. coli* DNA-directed DNA polymerase I, for example, can use RNA as a template, although there is a concomitant ~100-fold increase in dNTP K_m (Ricchetti and Buc, 1993). T7 DNA-directed RNA polymerase can also use RNA as a template (Konarska and Sharp, 1989). These are not exceptional observations

because it is a general property of polymerases that they display relaxed template specificity, at least *in vitro*.

While template specificity may be relaxed, polymerase substrate specificity is normally extremely stringent. T7 5 DNAP, for example, displays at least 2,000-fold selectivity for dNTPs over rNTPs, even in Mn⁺⁺ buffer which relaxes the ability of the polymerase to discriminate between dNTPs and ddNTPs (Tabor and Richardson, 1989).

It has been reported that transcripts synthesized by a 10 T7 RNAP Y639F mutant *in vivo* yielded 1/2-1/3 of the protein per transcript compared to transcripts synthesized by the wild-type enzyme (Makarova, *et al.*, 1995). The latter phenotype was unique to the Y639F mutant amongst a number of other active site mutants examined for *in vivo* expression, 15 and indicated that Y639F transcripts contained a defect that led to their being inefficiently translated.

A polymerase with an altered substrate specificity would be useful in many molecular biological applications, such as creating a nucleic acid molecule comprising at least 20 one non-canonical nucleotide.

Summary of the Invention

We disclose herein the identification of mutant polymerases, such as T7-type RNAPs, that display the ability to use dNTPs. The mutations occur in tyrosine 639 within 25 motif B (Delarue, *et al.*, 1990) of T7 RNAP.

We have characterized the ability of the Y639 mutants, as well as a large number of other active site mutants, to use dNTPs in both Mg⁺⁺ and Mn⁺⁺ buffers. Our results point to a specialized role for tyrosine 639 in T7 RNAP--and the 30 corresponding amino acid in other polymerases--in insuring that substrates to be added to the growing nucleic acid have

CONFIDENTIAL

the correct structure. The results reveal that both transcript and substrate structure affect the efficiency with which the transcript is extended and show that the restriction of unprimed initiation to RNA polymerases is not due to an intrinsic property of ribo- vs. deoxynucleotides, but simply to the selectivity of the polymerase active site. The present invention provides researchers with novel polymerase reagents and improved methods that expand the structural range of nucleic acids that can be enzymatically synthesized *in vitro*.

The present invention requires a polymerase with a reduced discrimination between canonical and non-canonical nucleoside triphosphates. In a preferred embodiment of the present invention, the polymerase has a reduced discrimination between rNTPs and dNTPs. In an especially preferred embodiment, the reduced discrimination is at least 10-fold compared to wild-type enzymes.

In one embodiment, the present invention is a method for synthesizing a nucleic acid molecule that comprises at least one non-canonical nucleotide. This method comprises the steps of incubating a template nucleic acid in a reaction mixture suitable for nucleic acid polymerization containing a mutant nucleic acid polymerase and the appropriate canonical and non-canonical nucleoside triphosphates which are substrates for a mutant nucleic acid polymerase and which are desired to be incorporated into the synthesized nucleic acid molecule.

In an especially preferred form of this method, the synthesized nucleic acid molecule has an altered susceptibility to a nuclease compared to a nucleic acid which could be synthesized using the corresponding non-

DRAFT
02/26/2024

mutant nucleic acid polymerase with canonical nucleoside triphosphates.

The present invention is also a method for determining the sequence of a nucleic acid molecule using a mutant RNA 5 polymerase.

The method comprises synthesizing a nucleic acid molecule, either *de novo* from a promoter, or by extending a primer annealed to the template molecule in four separate reactions. The four separate reactions each have all 4 10 rNTPs and a portion of a ddNTP, or have all 4 dNTPs and a portion of a ddNTP, or have 4 2'-fluorine-substituted NTPs and a portion of a ddNTP. Chain termination will occur and the products may be evaluated so that the sequence of the template molecule may be deduced. In one embodiment of this 15 method, the reactions which include a ddNTP occur in the same reaction mixture and are linked to a method for nucleic acid amplification, including, but not limited to, NASBA, 3SR, TMA, or other similar methods.

The present invention is also a partial ribo- 20 substitution method for determining the sequence of a nucleic acid molecule. This method comprises synthesizing a nucleic acid molecule, either *de novo* from a promoter or by extending a primer annealed to the template molecule in four separate reactions. The reactions each have, either four 25 dNTPs and a portion of an rNTP or four 2'-F-NTPs and a portion of an rNTP, or four different non-canonical nucleoside triphosphates, wherein these nucleoside triphosphates have substituents different than a hydroxyl group at the 2' position of the ribose and which the mutant 30 polymerase can use as substrates for synthesis in nucleic acids, and a portion of an rNTP. The reaction products are then cleaved at sites containing an incorporated rNTP by

using an alkaline solution or an RNase, and the cleaved nucleic acid fragments are separated according to size so that the sequence of the template molecule may be determined.

5 The present invention is also embodiments of a partial ribo-substitution method wherein the nucleic acid synthesis reactions of said method occur in the same reaction mixture and are also part of or linked to a method for nucleic acid amplification, including, but not limited to, NASBA, 3SR, 10 TMA, or other similar methods.

In still other embodiments of the present invention, the products of either 1, 2, 3, or 4 of the di α oxy-sequencing reactions or of the partial ribo-substitution sequencing reactions are performed or analyzed to determine 15 the presence or absence of a particular nucleic acid, or its relatedness to another nucleic acid, or whether it contains a mutation compared to another nucleic acid.

The present invention is also a kit for performing any of the above-identified methods.

20 It is an object of the present invention to provide a mutant polymerase capable of altered discrimination between canonical and non-canonical nucleoside triphosphates.

It is an object of the present invention to provide an improved DNA sequencing method.

25 It is an object of the present invention to provide a method to detect the presence of a nucleic acid.

It is an object of the present invention to provide a method to detect the identity of a nucleic acid.

30 It is an object of the present invention to provide a method to detect mutations in a nucleic acid.

It is an object of the present invention to minimize the steps involved in amplifying and sequencing, detecting, identifying and detecting mutations in nucleic acids.

It is another object of the present invention to provide a method for synthesizing nucleic acid molecules with altered nuclease susceptibility.

It is another object of the present invention to provide a method for synthesizing nucleic acid molecules comprising at least one non-canonical nucleoside triphosphate.

Other objects, features and advantages of the present invention will become apparent after examination of the specification, claims and drawings.

Description of the Drawings

Fig. 1 diagrams the transcription products produced by Y639F and w.t. T7 RNAP in the presence of various combinations of rNTPs and dNTPs.

Fig. 2 shows the effect of dGTP substitution on transcription by the w.t. and Y639F polymerase.

Fig. 3 shows the effects of single-strandedness and sequence in the initially transcribed region on the activity of Y639F in reactions with 4 rNTPs or 4 dNTPs.

Fig. 4 shows transcription by Y639F and w.t. polymerase with dGTP or rGTP on poly(dI)•poly (dC).

Fig. 5 shows primed synthesis of DNA and RNA with Y639F and the w.t. polymerase.

Fig. 6 shows relative elongation rates of Y639F in "4 rNTP" and "3 rNTP + 1 dNTP" reactions.

Fig. 7 is a diagram of the mutagenesis strategy involved in creating a mutant SP6 RNA polymerase.

Description of the Invention

Definitions

By "mutant polymerase" is meant a nucleic acid polymerase which has at least one altered amino acid compared to the corresponding wild-type polymerase, wherein said mutation or alteration results in the mutant polymerase having a reduced discrimination between non-canonical and canonical nucleoside triphosphates as substrates.

By "template" we mean a macromolecular pattern or mold for the synthesis of another macromolecule, composed of a sequence of nucleotides, either rNTPs or dNTPs, that serves to specify the nucleotide sequence of another structure.

"Nucleotide" refers to a base-sugar-phosphate compound. Nucleotides are the monomeric subunits of both types of nucleic acid polymers, RNA and DNA. "Nucleotide" refers to ribonucleoside triphosphates, rATP, rGTP, rUTP and rCTP, and deoxyribonucleoside triphosphates, such as dATP, dGTP, dTTP, dCTP.

As used herein, "nucleoside" refers to a base-sugar combination without a phosphate group. "Base" refers to the nitrogen-containing base, for example adenine (A), cytidine (C), guanine (G) and thymine (T) and uracil (U).

"Incorporation" refers to becoming a part of a nucleic acid polymer. There is a known flexibility in the terminology about incorporation of nucleic acid precursors. For example, the nucleotide dGTP is a deoxyribonucleoside triphosphate. Upon incorporation into DNA, it becomes a dGMP, or deoxyguanosine monophosphate moiety. Although there is no dGTP molecule in DNA, one may say that one incorporates dGTP into DNA.

As defined herein, a "canonical" nucleoside triphosphate for an RNA polymerase ("RNAP") consists of any

000000000000000000000000

ribonucleoside-5'-triphosphate ("rNTP" or "NTP") which has an hydroxyl group at the 2'-position of the sugar, including, but not limited to, the four common ribose-containing substrates for an RNA polymerase -ATP, CTP, GTP and UTP. A 2'-deoxyribonucleoside-5'-triphosphate ("dNTP") which has hydrogen at the 2'-position of the sugar, including, but not limited to, the four common deoxyribose-containing substrates (dATP, dCTP, dGTP and dTTP, also known as "TTP") for a DNA polymerase ("DNAP") is defined herein as a "non-canonical" nucleoside-5'-triphosphate or a "non-canonical NTP" or a "non-canonical nucleotide" or a "non-canonical deoxynucleotide" or a "non-canonical triphosphate" or a "non-canonical substrate" for an RNA polymerase. On the other hand, a "canonical" nucleoside triphosphate for a DNAP consists of any dNTP which has a hydrogen at the 2'-position of the sugar, while an rNTP is defined as a "non-canonical NTP" or a "non-canonical nucleotide" or a "non-canonical substrate" for a DNAP. The terms "canonical" and "non-canonical" are meant to be used herein only with reference to the 2' position of the sugar. Thus, as defined herein, 2',3'-dideoxynucleoside-5'-triphosphates ("2',3'-ddNTPs" or "ddNTPs") are "non-canonical" substrates for an RNAP, but are defined as "canonical" for a DNAP. Further, any other substituent than an hydroxyl group at the 2'-position of ribose or a hydrogen at the 2'-position of deoxyribose, including, but not limited to, a fluorine ("F" or "fluoro" group) or an amino group, would be defined as "non-canonical" for both RNAPs and DNAPs herein. The terms "canonical" or "non-canonical" also are not meant to refer to the nucleic acid bases attached to the sugar moieties. Thus, for example, other natural or modified nucleic acid

bases attached to the 1'-position of ribose-5'-triphosphate would still be defined as "canonical" herein.

By "a (mutant) nucleic acid polymerase (enzyme) with reduced discrimination between canonical and non-canonical nucleoside triphosphate substrates", we have a specific quantitative definition calculated as follows:

1. One first determines the K_m and the k_{cat} for each enzyme (mutant and wild-type) using the non-canonical nucleotide as a substrate and using the canonical nucleotide as a substrate, as was described previously (Patra, et al., 1992). The value " K_m " expresses how readily the enzyme will bind the substrate (a larger K_m implies weaker binding) and " k_{cat} " expresses the rapidity with which the substrate, once bound by the enzyme, is reacted upon.
2. One next calculates the numerical value for k_{cat}/K_m for each enzyme and for each substrate. By broad scientific consensus, the specificity of an enzyme for a substrate is felt to be most suitably expressed by this ratio.
3. For each enzyme, one then calculates the numerical value obtained by using the value of k_{cat}/K_m for the canonical substrate in the numerator and k_{cat}/K_m for the non-canonical substrate in the denominator. This number indicates how much a given enzyme discriminates between the two substrates. For example, if this value equals 1, then the enzyme uses both the canonical and the non-canonical substrates equally well; it does not discriminate between the two substrates. If this value is greater than 1, then the enzyme discriminates by that factor between the two substrates; for example, if the value is 100, then the enzyme discriminates by a

00000000000000000000000000000000

factor of 100 in favor of the canonical substrate compared to the non-canonical substrate. Similarly, a value less than 1 means that the enzyme discriminates in favor of the non-canonical substrate over the canonical substrate.

5 4. Finally, one compares the numerical value calculated in step 3 above for the wild-type (w.t.) enzyme with the value calculated in step 3 for the mutant enzyme. If the value calculated for the mutant enzyme is less than the value calculated for the wild-type enzyme, then the 10 mutant enzyme "has a reduced discrimination for the non-canonical substrate compared to the wild-type enzyme." For example, if the values calculated in step 3 are 100 for the wild-type enzyme and 10 for the 15 mutant enzyme, then the discrimination of the mutant enzyme in favor of the canonical substrate (or against the non-canonical substrate) is reduced 10-fold compared to the discrimination of the wild-type enzyme.

We have found that for wild-type T7 RNAP, the average 20 of the k_{cat}/K_m values for the four common rNTPs (ATP, CTP, GTP & UTP) is about 120-fold larger than the average k_{cat}/K_m values for the four common dNTPs (dATP, dCTP, dGTP and dTTP); i.e., the wild-type enzyme discriminates by a factor of 120 for rNTPs vs. dNTPs. For the Y639F mutant enzyme, 25 the average of the k_{cat}/K_m values for the four common rNTPs is only about 6-fold larger than the average k_{cat}/K_m values for the four common dNTPs. Thus, using the average k_{cat}/K_m values for these substrates, the Y639F mutant T7 RNAP enzyme has about a 20-fold reduced discrimination between dNTPs and 30 rNTPs. However, it is recognized that the difference in discrimination between wild-type and mutant enzymes will vary depending on the non-canonical substrates and the

mutant enzymes used. Therefore, for the purposes of this invention, we herein define "a (mutant) nucleic acid polymerase (enzyme) with reduced discrimination between canonical and non-canonical nucleoside triphosphate substrates" as "a polymerase which has at least a 10-fold reduced discrimination compared to the corresponding wild-type enzyme for non-canonical nucleotides compared to canonical nucleotides, wherein the respective values for discrimination between canonical and non-canonical substrates is calculated using the average of the k_{cat}/K_m values for all four rNTPs and all four dNTPs."

By "T7-type RNA polymerases" we mean T7, T3, φI, φIIH, W31, ghl, Y, A1122, SP6 and mitochondrial RNAPs.

In General

There are many reasons to synthesize nucleic acid molecules containing at least one non-canonical nucleotide. For example, incorporation of a non-canonical nucleotide may make the synthetic nucleic acid more resistant and therefore, more stable, to a nuclease, such as a ribonuclease. Also, one may wish to incorporate one or more non-canonical nucleotides which, for example, change the nuclease digestion pattern so that the product nucleic acid is easier to detect or characterize. For example, because RNase A cleaves RNA only after C or U, replacement of one or both of these rNMPs by a dNMP or other non-canonical nucleotide that is resistant to cleavage by RNase A would alter the digestion pattern of the nucleic acid.

There are many uses for nucleic acids which have one or more of these properties, such as nuclease resistance. For example, nucleic acids containing at least one non-canonical nucleotide may have advantages for use as ribozymes, or as

00000000000000000000000000000000

nucleic acid molecules used for gene therapy, in a vaccine,
in an antiviral composition, in an antimicrobial
composition, in an anti-sense composition for regulating
gene expression, in a composition for hybridization to a
complementary nucleic acid, including as a primer, or as a
probe for detection of a complementary nucleic acid for a
variety of purposes.

Some nucleic acid molecules, such as those of mixed
dNMP/rNMP composition, are highly useful for certain
applications, but are presently difficult or impossible to
produce on a practical scale. Thus, improved methods for
synthesizing such nucleic acid molecules *in vitro* would be
highly desirable. For example, probes of mixed DNA-RNA-DNA
composition for the Cycling Probe Assay (Duck, P.G., *et al.*,
1990) are currently made using difficult chemical methods.

We describe herein previously-unknown properties of T7-type RNA polymerases having a non-wild-type amino acid at specific positions within the polypeptides. We found that altering the amino acid at these specific positions results in mutant polymerases having at least a 10-fold reduced discrimination between 2'-deoxyribonucleoside-5'-triphosphates (dNTPs) and ribonucleoside-5'-triphosphates (rNTPs) as substrates in *in vitro* nucleic acid synthesis reactions compared to the corresponding wild-type enzymes.
We found that these mutant polymerases also have reduced discrimination for other non-canonical nucleoside triphosphate (NTP) substrates, including 2',3'-dideoxy-ribonucleoside-5'-triphosphates (ddNTPs) and 2'-fluoro-nucleoside-5'-triphosphates (2'-F-NTPs). Based on knowledge of these novel properties, we have disclosed methods for using these mutant polymerases for producing nucleic acid molecules containing at least one non-canonical nucleotide

and for characterizing nucleic acid molecules by synthesizing nucleic acid molecules containing at least one non-canonical nucleotide *in vitro*.

In one preferred embodiment, the invention uses mutant RNA polymerases that efficiently utilize deoxynucleoside triphosphates as substrates. *In vitro*, this mutant will synthesize RNA, DNA, or 'transcripts' of mixed dNMP/rNMP composition from a template molecule depending on the mix of NTPs or dNTPs present in the synthesis reaction.

10 Mutant Polymerases of the Present Invention

In a preferred embodiment, the polymerase mutation is conservative, for example, changing tyrosine 639 (of T7 polymerase) within the active site to phenylalanine, and does not substantially affect promoter specificity or overall activity. Non-conservative mutations of this tyrosine also reduce discrimination between deoxy- and ribo-nucleoside triphosphates, but these mutations also typically cause large activity reductions. Among the most active of the non-conservative mutations, enzymes with methionine or leucine in place of the wild-type tyrosine at the 639 position of T7 RNAP had about half the enzymatic activity of the wild-type enzyme.

Of 26 other mutations of other amino acid positions examined in and around the active site of T7 RNAP, none showed marked effects on rNTP/dNTP discrimination.

T7 RNA polymerase can use RNA templates as well as DNA templates and is capable of both primer extension and *de novo* initiation. The Y639F mutant, described below in the Examples, retains the ability to use RNA or DNA templates. Thus, this mutant can display *de novo* initiated or primed DNA directed DNA polymerase, reverse transcriptase, RNA directed RNA polymerase, or DNA directed RNA polymerase

activities depending on the templates and substrates presented to it in the synthesis reaction.

A major theme of research on nucleic acid polymerases over the past several years has been the discovery of extensive structural similarity between the majority of these enzymes, even those from functionally different classes (Sousa, *et al.*, 1993; Pelletier, *et al.*, 1994; Jacob-Molina, *et al.*, 1993; Kohlstaedt, *et al.*, 1992; Sawaya, *et al.*, 1994; Steitz, *et al.*, 1994). One part of this work has been the identification of well-conserved residues. Five amino acids have been identified as invariant in a large number of DNA-directed RNA polymerases (Delarue, *et al.*, 1990). In T7 RNAP these are D537, K631, Y639, G640A and D812. A specific, conserved function has been revealed for the two invariant aspartates in coordinating the catalytic Mg⁺⁺ ion (Sousa, *et al.*, 1993; Pelletier, *et al.*, 1994; Jacob-Molina, *et al.*, 1993; Kohlstaedt, *et al.*, 1993; Sawaya, *et al.*, 1994; Steitz, *et al.*, 1994). Our observations imply a similarly specific and conserved function for Y639 as a sensor of inappropriate geometry or structure in the template-NTP-primer/RNA complex.

The present invention encompasses methods for synthesis of nucleic acids containing at least one non-canonical nucleotide using mutant nucleic acid polymerases which have reduced discrimination for non-canonical nucleoside triphosphate substrates. The examples below demonstrate the reduced dNTP/rNTP discrimination of mutants of T7 RNA polymerase and SP6 RNA polymerase. Genes for other polymerases may be modified or mutated to obtain mutant enzymes which have similar reduced discrimination for non-canonical substrates. If one wished to mutate an RNA

polymerase to have the properties described herein, one would first locate the amino acid corresponding to the T7 polymerase Y639 in other RNA polymerases. Identification of the corresponding mutation site in other polymerases can be
5 done by the well-established procedure of sequence alignment, which involves aligning the amino acid sequences of two proteins, introducing gaps and insertions, and shifting the sequences with respect to each other while maintaining their original linearity. Such alignment procedures are often performed with the aid of one or more
10 computer programs into which the amino acid sequences that one wishes to compare have been entered. When the sequence identity of two proteins is high enough (greater than or equal to 30%) over a sufficient length of amino acids
15 (greater than or equal to 50), this procedure is very reliable in identifying amino acids that occupy corresponding structural and functional positions in the two proteins. Such conditions are met for the T7-type group of RNAPs, which include T7, T3, ϕ I, ϕ IIH, W31, ghl, Y, A1122,
20 SP6 and mitochondrial RNAPs, and allow identification of the mutation site corresponding to Y639 in T7 RNAP.

Using this method, we predicted that the amino acid in w.t. SP6 RNAP that corresponded to the Y639 site in T7 RNAP was Y631, and as described herein, mutagenesis of this site
25 resulted in a Y631F mutant SP6 RNAP which has a similar reduced discrimination for dNTPs compared to rNTPs like the Y639F mutant T7 RNAP.

From alignment studies, it is known that there is a conserved motif present in T7-like RNAPs and class I DNAPs
30 with the following consensus sequences:

... K - - - - - - - Y G ...

DRAFTS 6
DATE 10/2

where Y is the tyrosine at amino acid number 639 in the T7 RNA polymerase protein. The same consensus sequence is observed in the SP6 RNA polymerase and T3 RNA polymerase proteins where a K (K=lysine) is succeeded by 7 amino acids and a Y G (G=glycine). In SP6 RNA polymerase the Y is at amino acid number 631 in the polypeptide chain, and in T3 RNA polymerase it is at amino acid number 573. By mutating the codon for Y631 in SP6 RNA polymerase such that a phenylalanine is at this position, the expected phenotypic change was realized.

In summary, one may locate the corresponding mutation site in other RNAPs by aligning the amino acid sequence of a T7-like RNAP, chosen from among T7, T3, φI, φIIH, W31, ghl, Y, A1122, SP6 and mitochondrial RNAPs, against the conserved motif given above and identifying which position corresponds to the Y639 position in T7 RNAP.

As stated above, the conserved motif is also present in class I DNAPs. While a structure of T7 RNAP complexed with NTP is not available, the structure of the homologous Klenow fragment of DNAP I with dNTP has been obtained (Beese, et al., 1993). This structure demonstrated that the amino acids encompassed within the above-mentioned conserved motif (i.e., amino acid residues 758 to 767 of *E. coli* DNAP I) are in proximity to the deoxyribose sugar of the dNTP, so that it is reasonable that mutations within this motif might affect the ability of a DNAP to discriminate between dNTPs and rNTPs. However, one of the present inventors found that when a mutation was made which changed the amino acid at the position in a class I DNAP corresponding to the Y639 mutation in T7 RNAP, the mutant DNAP retained enzymatic activity, but did not have a reduced discrimination for rNTPs compared to dNTPs. Thus, it is not obvious what

mutations, if any, would result in a class I DNAP having reduced discrimination for rNTPs vs. dNTPs, even if it is reasonable to assume that such a mutation would occur within the above-mentioned conserved motif (residues 758 to 767 of 5 *E. coli* DNAP I) which the structure shows to be in proximity to the dNTP.

Because the structure of the Klenow fragment of DNAP I complexed with dNTP was determined (Beese, et al., 1993), researchers have believed that the homologous conserved 10 motif of T7 RNAP (i.e., amino acid residues 631-640) is likely to be in proximity to the ribose moiety of an NTP, as the case for the DNAP I. Nevertheless, prior to the work of the present invention, it was not possible to know which mutation, if any, might result in a reduced discrimination 15 for dNTPs vs. rNTPs. Since the Y639 mutation of T7 RNAP was identified, as presented herein, one of the present inventors has modeled NTP in T7 RNAP (Huang, et al., submitted for publication) based on the structures of the homologous Klenow fragment of DNAP I complexed with dNTP 20 (Beese, et al., 1993) and of RT complexed with primer-template (Jacabo-Molina, et al., 1993). Models based on either structure agree in placing the ribose close to Y639, and in revealing no other side chain capable of discriminating the hydrogen bonding character of the 2'- 25 substituent within 5 angstroms of the 2'-group of the NTP. Thus, the model is consistent with our results related to a reduced discrimination of Y639 RNAP mutants for dNTPs vs. rNTPs, even though additional studies (Huang, et al., submitted for publication) have determined that the hydrogen 30 bonding character is not the only factor involved in dNTP/rNTP discrimination.

DRAFT

Less is known about non-T7-type RNAPs. For many, the amino acid sequences are not known. Non-phage-encoded host bacterial RNA polymerases are complex multi-subunit proteins. A nucleotide polymerization site has been 5 localized in the β subunit of *E. coli* RNA polymerase although participation of other subunits is not ruled out. In order to determine the site in a non-T7-like RNAP which would result in a reduced dNTP/rNTP discrimination, one would first use the above-described procedure of alignment 10 to determine if the

... K — — — — Y G ... motif was present. If so, it may be possible to obtain the desired mutation in the same manner as for T7-like RNAPs. However, if the conserved motif is not present, one may obtain the desired mutation 15 with greater difficulty by random mutagenesis and enzyme assay screening in order to find a change or changes that result in reduced dNTP/rNTP discrimination.

Once one has determined where the corresponding Y639 site is in the polymerase one wishes to mutate, one would 20 use standard methods in the art of molecular biology to create an amino acid substitution. As disclosed above, a conservative substitution is preferable. For example, a substitution of a phenylalanine for a tyrosine is most preferable. The Examples below disclose a method for 25 creating a mutant polymerase, but one of skill in the art will realize that there are many substitute methods of equal effectiveness.

Methods of the Present Invention

In one embodiment, the present invention is a method 30 for using a mutant polymerase for synthesizing *in vitro* a nucleic acid molecule which comprises at least one non-

00065346 000102

canonical nucleotide in place of at least a portion of the canonical nucleotides. The method comprises the steps of incubating a template nucleic acid in a reaction mixture containing a mutant nucleic acid polymerase which has reduced discrimination between canonical and non-canonical nucleoside triphosphates, including between dNTPs and rNTPs, and the appropriate canonical or non-canonical nucleoside triphosphates which are substrates for the nucleic acid polymerase. One then follows standard polymerase reaction protocols and creates the synthesized nucleic acid molecule.

Preferably, the reactions also contain inorganic pyrophosphatase, which is known to increase the yields in *in vitro* transcription reactions (Cunningham, P.R. and Ofengand, J., 1990) and to reduce pyrophosphorolysis in *in vitro* DNA synthesis reactions (Tabor, S., and Richardson, C.C., 1990), as well as buffers and other components which are known to those of skill in the art to be optimal for the particular w.t. polymerase used. Cunningham and Ofengand (1990) provide an example of conditions which may be used for unprimed synthesis with T7 RNA polymerase or mutant T7 RNAPs, although one of skill in the art will recognize, with respect to reactions with these enzymes or other enzymes, the need to optimize the concentrations and ratios of canonical and non-canonical NTP substrates according to the respective K_m and application and to modify reaction conditions, such as temperature, amount of enzyme, salt concentration, or divalent cation (e.g., Mg²⁺ or Mn²⁺) concentration, in order to produce improved results such as higher yield or a greater percentage of full-length products.

In a preferred form of this method, the resulting synthesized nucleic acid molecule has a different

DRAFT
09052345 0412102

susceptibility to a nuclease compared to a nucleic acid synthesized by the corresponding non-mutant nucleic acid polymerase under identical reaction conditions with canonical substrates. By "different susceptibility" we mean 5 to include reduced, increased, or, in the case of synthetic nucleic acids containing both canonical and non-canonical nucleotides, altered susceptibility to a nuclease, which may be either a DNase or RNase. The nature of the reduced, increased or altered susceptibility to a nuclease is also 10 related to the properties of the nuclease. For example, a nucleic acid resistant to RNase A, which cleaves RNA only after C or U, may be synthesized using fewer non-canonical nucleotides (e.g., dNTPs or 2'-F-NTPs) than a nucleic acid which is resistant to RNase I, which cleaves after every 15 base.

In a preferred form of the present invention, the resulting synthesized nucleic acid is a ribozyme or a nucleic acid molecule used for gene therapy, in a vaccine as an antiviral composition, in an antimicrobial composition, 20 as an antisense composition for regulating gene expression, in a composition for hybridization to a complementary nucleic acid, such as for a primer, or as a probe for detection of a complementary nucleic acid.

The resulting synthesized nucleic acid may be either 25 single- or double-stranded.

The present invention is also a kit for performing the method of synthesizing a nucleic acid containing at least one non-canonical nucleotide. Typically, the kit contains a mutant nucleic acid polymerase with a reduced discrimination 30 for non-canonical compared to canonical substrates and data or information describing conditions under which the method may be performed.

The present invention is also improved methods for sequencing nucleic acids using a mutant nucleic acid polymerase of the present invention.

Because 2',3'-dideoxynucleotides are not substrates for wild-type RNA polymerase, it previously has not been possible to use the Sanger method for determining the sequence of a nucleic acid with an RNA polymerase, although 3'-deoxy- or 3'-hydroxymethyl analogs have been used as terminators for Sanger-like sequencing with RNA polymerases.

However, 2',3'-ddNTPs are substrates for the mutant nucleic acid polymerases of this invention which can also utilize both rNTPs and dNTPs as substrates, and the present invention is also a method for sequencing nucleic acid molecules (DNA or RNA) using a mutant nucleic acid polymerase and 2',3'-ddNTPs as terminators.

In one embodiment of this method, the nucleic acid to be sequenced, whether DNA or RNA, is used as a template for *in vitro* nucleic acid synthesis from a primer (i.e., primed synthesis) using a mutant RNA polymerase which has a reduced discrimination for dNTPs compared to rNTPs. Each of four different reactions also contains an amount of at least one nucleoside triphosphate corresponding to each nucleic acid base represented in either DNA or RNA, chosen from among the 2'-deoxynucleotides dATP, dCTP, dGTP and dTTP or dUTP, or the four common ribonucleotides ATP, CTP, GTP and UTP, or the 2'-fluorine-substituted nucleotides 2'-F-ATP, 2'-F-CTP, 2'-F-GTP and 2'-F-UTP or 2'-F-TTP. A 2',3'-dideoxynucleotide is also included in each *in vitro* nucleic acid synthesis reaction in an amount that will result in random substitution by the dideoxynucleotide of a small percentage of the corresponding rNTP, dNTP or 2'-F-NTP that is present

in the reaction and that would be incorporated into the synthetic nucleic acid in a template-dependent fashion.

Thus, each DNA synthesis reaction yields a mixture of DNA fragments of different lengths corresponding to chain termination wherever the dideoxynucleotide was incorporated in place of the ribo-, 2'-deoxy- or 2'-fluoro- nucleotide. The DNA fragments are labelled, either radioactively or non-radioactively, by one of several methods known in the art and the label(s) may be incorporated into the DNA by extension of a labelled primer, or by incorporation of a labelled ribo-, deoxy-, 2'-fluoro- or 2',3'-dideoxy-nucleotide.

By carrying out DNA synthesis reactions for each of the dideoxynucleotides (ddATP, ddCTP, ddGTP and ddTTP or ddUTP), then separating the products of each reaction in adjacent lanes of a denaturing polyacrylamide gel or in the same lane of the gel if different distinguishable labels are used for each separate reaction, and then detecting those products by one of several radioactive or non-radioactive methods known in the art, the sequence of the DNA template can be read directly. Also, other matrices than polyacrylamide which separate the fragments based on size may be used. Those with skill in the art will also recognize that other nucleotide analogs generally used for reducing sequencing compressions, such as ribo-, deoxy- or 2'-fluoro- nucleoside triphosphates containing 7-deaza-guanine or inosine, may also be used in place of the ribo-, deoxy- or 2'-fluoro-nucleotide for which the respective analog is used.

The present invention also comprises another embodiment of this method for sequencing using 2',3'-ddNTPs and a mutant RNA polymerase which has a reduced discrimination for dNTPs compared to rNTPs. In this embodiment of the method,

the nucleic acid to be sequenced, whether DNA or RNA, is used as a template for *de novo* (i.e., unprimed) *in vitro* nucleic acid synthesis beginning at an RNA polymerase promoter. In this embodiment, a primer is omitted from the reactions and depending on the promoter sequence, in addition to the other components used in the first embodiment, an amount of a dinucleoside tetraphosphate or a trinucleoside penta-phosphate may be added as an initiating nucleotide (Moroney and Piccirilli, 1991) so that the majority of nucleic acid synthesis is initiated from a single site. Because no primer is used, the labelling of the sequencing products must be carried out by one of the other methods envisioned and discussed with respect to the first embodiment of the method or by incorporating a label into or on an initiating dinucleotide or trinucleotide. In other respects, the second embodiment of this sequencing method of the invention is similar to the first embodiment of the method.

The present invention also comprises methods of sequencing nucleic acids by partial ribo-substitution using mutant nucleic acid polymerase which have a reduced discrimination between non-canonical and canonical nucleotides. These methods have advantages over the partial ribo-substitution sequencing method described by Barnes (1977), which relied on the use of a Mn²⁺-containing reaction buffer to relax the ability of a wild-type DNA polymerase to discriminate between dNTPs and rNTPs. Further, only DNA was used as a template for this method and *de novo* (i.e., unprimed) nucleic acid synthesis was not envisioned. In contrast, the ribo-substitution sequencing method of the present invention uses a mutant nucleic acid polymerase which has an inherent reduced discrimination

between dNTPs and rNTPs and, although it may be included, Mn²⁺ is not required in the sequencing reactions. In still another embodiment, 2'-fluorine-substituted NTPs are used in place of dNTPs in the ribo-substitution reaction.

5 Embodiments of the method of the present invention also include sequencing using either DNA or RNA as templates and sequencing using either nucleic acid primers or de novo nucleic acid synthesis from an RNA polymerase promoter sequence.

10 Since incorporation of the sequence-delimiting ribonucleotide does not terminate nucleic acid synthesis during partial ribo-substitution sequencing, all of the radioactive or non-radioactive label must be incorporated into the sequencing reaction products prior to incorporation 15 of the first ribonucleotide in order to avoid multiple labeled produced from each nucleic acid molecule synthesized. Multiple labeled products, starting from different positions on the template, would make it difficult or impossible to interpret sequence results. Thus, labeling 20 of nucleic acid products for partial ribo-substitution sequencing must be accomplished by means such as incorporating the label into a primer, when used, or into an initiating di- or tri- nucleotide, or by prelabelling in the presence of an amount of a labeled nucleotide which will be 25 used up or destroyed and/or limiting 2'-deoxy- or 2'-fluoro-nucleotides prior to addition of the sequence-delimiting rNTPs.

We envision sequencing reactions wherein all of the distinguishable non-radioactive labels in, or attached to, 30 or connected with, the products from more than one of the reactions, up to and including all four of the sequencing reactions for a template, or even reactions from multiple

templates if distinguishable non-radioactive labels are present, are detected in or from a single lane of a sequencing gel or a single capillary electrophoresis tube or a single matrix or means of any kind using an automated sequencer or other detection device.

Examples

Example 1: Creation and Characterization of a Mutant T7 RNA Polymerase

A. Materials and Methods

Nucleic acids and NTPs: Nucleotides were from 10 Pharmacia or USB/Amersham. Polynucleotides were from Pharmacia and the Midland Certified Reagent Company. Synthetic DNAs were prepared at the UTHSCSA DNA synthesis facility on an Applied Biosystems DNA synthesizer and purified by HPLC. A synthetic RNA 12mer was from the 15 Midland Certified Reagent Company. Plasmids pT75 (Tabor and Richardson, 1985) and pBS (Stratagene Inc.) were purified from *E. coli* by alkaline-lysis and cesium chloride gradient centrifugation (Sambrook, *et al.*, 1989). Radioactive 20 nucleotides were from NEN Dupont or ICN.

Preparation and purification of mutant polymerases: Construction, expression, and purification of the T7 RNAP mutants was described previously (Bonner, *et al.*, 1992).

Transcription reactions: Transcription reactions were carried out in 40 mM Tris-HCl pH 8.0, 15 mM MgCl₂, and 5 mM DTT or 20 mM Manganese Citrate pH 8.0, 5 mM DTT at 37°C. Template, polymerase, and NTP concentrations were as indicated in the legends to the figures and tables. Relative activity determinations were made by taking 4 µl aliquots of reactions at 5, 10, and 20 minute time points 30 and spotting on to DE81 filter paper. Unincorporated

DECODED
DRAFT

nucleotides were separated from incorporated nucleotides by washing the filter paper with .5 M KH₂PO₄, pH 7.0 and retained radioactive nucleotide was quantitated with a Molecular Dynamics phosphorimager. Radioactive NTPs used 5 were as indicated in figure and table legends. To evaluate rNTP/dNTP selectivity, reactions were run with 4 rNTPs and pT75 as template and α -P³² rNTPs or α -P³² dNTPs were used to label the transcripts. The relative rate of incorporation 10 of an rNTP vs. its cognate dNTP was determined from the relative percentages of labeled rNTP vs. dNTP incorporated into DE81 retainable RNA at 5, 10, and 20 minute time points. Apparent miscoding frequencies were determined similarly, though in this instance the template was a single-stranded homopolymer and the reaction contained the 15 complementary unlabeled rNTP, and complementary α -P³² rNTP or one of the 3 non-complementary α -P³² rNTPs. The relative percentages of labeled complementary vs. non-complementary rNTPs incorporated at 5, 10 and 20 minute time points gave an apparent miscoding rate. The rNTP/dNTP selectivity assay 20 was used to test the following T7 RNAP mutants for effects on substrate discrimination: D537S, D537E, S539A, R551S, D552S, R627S, K631S, L637A, Y639S, Y639F, Y639A, G640A, F644A, G645A, Q649S, I810S, H811S, H811A, D812A, D812E, D812S, D812N, D879E+ Δ F882+ Δ A883, D879E+A881T+ Δ A883, 25 D879E+F884+A885, D879E+F882Y, D879E+ Δ A883, D879E+F882W, D879E.

Elongation rate determinations were carried out as described (Golomb & Chamberlin, 1974) with some variations (Bonner, et al., 1994).

30 Determination of NTP K_m and k_{cat} was as described previously (Patra, et al., 1992).

B. Results

Structure of the transcripts synthesized by Y639F and the w.t. enzyme with rNTPs and dNTPs: Fig. 1 diagrams the structure of transcription products produced by Y639F and w.t. T7 RNAP in the presence of various combinations of rNTPs and dNTPs. The template was pT75 (Tabor and Richardson, 1985) cut with *Hind*III so that transcription from its T7 promoter generates a 59-base run-off transcript. Electrophoresis was on a 20% polyacrylamide 6M urea gel.

Plasmid and polymerases were at concentrations of 10^{-7} M, and NTP concentrations were .5 mM (all rNTPs and dTTP), 1 mM (dATP, dGTP), or 5 mM (dCTP). γ -P³²-GTP was added to radiolabel the transcription products. Wild-type (WT) or Y639F mutant (Mu) polymerases and NTPs used are as indicated. Poly-rG products of various sizes are labeled in lane A ("2G", "3G", etc...) and heterogeneous sequence abortive transcripts of different lengths are indicated by "4H", "5H", etc... in lane c. Lanes q-t are a 10-fold longer exposure of lanes m-p.

Fig. 1 shows transcription reactions carried out with the w.t. enzyme or the Y639F mutant polymerase and a T7 ϕ -10 promoter template. Transcription by T7 RNA polymerase, like other RNA polymerases, is characterized by an initial, poorly processive 'abortive' phase of transcription during which the short, nascent transcript frequently dissociates from the ternary complex. When the transcript reaches a length of ~9 bases transcription becomes highly processive and the transcript becomes stably associated with the elongation complex. On this promoter, which initiates with GGGAGACCGGAAU (SEQ ID NO:1), T7 RNAP can also synthesize long poly-G ladders (labeled "2G", "3G", etc. in lanes a and b) when rGTP is the sole NTP present (Martin, *et al.*, 1988).

Lanes c and d of the electrophoretic gel display the transcription products typical of runoff reactions counting all 4 rNTPs: there are abortive transcripts ranging up to 8 bases in length and a 59-base runoff product. The 4mer
5 length the sequences of the poly-G transcripts and the abortive transcripts made in the presence of all 4 rNTPs (labeled "4H", "5H", etc.) diverge and no longer co-migrate with the poly-G transcripts of equivalent size.

When rATP is omitted from the transcription reactions
10 (lanes e and f) normal elongation of the initially synthesized GGG trimer cannot occur. There are no heterogenous sequence abortive transcripts or 59 base runoff products made and instead long poly-G transcript s, as in lane a, are made. Adding dATP to reactions lacking rATP
15 does not change the transcripts produced by the w.t. enzyme (lane g). However, with the mutant enzyme we observed that addition of dATP (lane h) allows synthesis a long runoff transcript as well as synthesis of heterogenous sequence abortive transcripts that do not co-migrate with the poly-G
20 transcripts. Observation of an abortive transcript in lane h running near the position of the "4H" band in lane d confirms extension of the GGG trimer with an A but note that the major 4mer transcript in lane h migrates close to, but not precisely with, the major 4mer in lane d or in adjacent
25 lane i. This is consistent with the expectation that these 4mers will have identical sequence and length but different structure (i.e., rGrGrGrA in lanes d or i; rGrGrGdA in lane h). It should also be noted that some poly-G transcript synthesis is observed in lane h. For example, in lane h we
30 observe both a heterogeneous sequence 4mer migrating near the "4H" position and a smaller amount of 4mer band migrating at the "4G" position. When 4 rNTPs are present

(lanes c or d) the synthesis of poly-G transcripts is more completely suppressed. This indicates that dATP is utilized by Y639F, but not as efficiently as rATP.

When rCTP is omitted from the reaction, transcripts 5 terminate predominately at the 6mer length because rCMP is normally first incorporated at position 7 (lanes i, j). Addition of dCTP does not allow extension of the 6mer in reactions with the w.t. enzyme (lane k). However, addition of dCTP to reactions with Y639F allows extension beyond the 10 6mer length and synthesis of the runoff transcript (lane l).

Again, the following should be noted: 1. The transcripts larger than 6 bases do not co-migrate with their counterparts in lanes c or d consistent with the expected structural difference despite length and sequence identity, 15 2. there is more termination at the 6- and 7mer points in lane l than in lanes c or d, indicating that Y639F uses dCTP well, but not as efficiently as it utilizes rCTP.

In lanes m and n UTP was omitted from the reactions. Lanes q-t show a 10-fold longer exposure of lanes m-p. 20 Within the set of 4 NTPs, UTP is unique on this template since it first becomes incorporated into the transcript at the 13 base position. This corresponds to a transcript length subsequent to the transition from abortive to processive transcription. As a consequence of this 25 transition, the ternary complex becomes more stable (Martin, et al., 1988). Therefore, when transcript extension is blocked during the processive phase of transcription, the stalled ternary complex does not rapidly dissociate (Shi, et al., 1988). Instead it remains stalled on the template, 30 near the promoter, and blocks reinitiation. For this reason we observe a large decrease in the overall amount of

transcription when UTP is omitted from the reaction in lanes m and n.

A longer exposure of lanes m and n (lanes q, r) does, however, reveal transcription products of the expected 5 structure. When TTP is added to these reactions (lanes o, p, s, t) synthesis of the 59 base runoff transcript is observed with both the w.t. and Y639F mutant.

The ability of the w.t. enzyme to extend transcripts with TTP when it is unable to extend transcripts with the 10 other dNTPs is also likely to be related to unique position at which UTP/TTP first becomes incorporated into the transcript. The amount of transcript termination or extension that occurs with a particular dNTP depends simply 15 on the relative rates of ternary complex dissociation or dNMP incorporation (McClure and Chow, 1980). During abortive transcription complex, dissociation must be more rapid than the rate at which the w.t. enzyme incorporates dNMPs into its transcripts. During processive transcription even the expected slow rate of dNMP incorporation by the 20 w.t. enzyme must be competitive with the slow rate of dissociation of the stable elongation complex, and an ability of the w.t. enzyme to incorporate dNMPs becomes manifest. Because elongation during the processive phase of transcription is fast (~230 bases/sec, Golomb and Chamberlin, 1974), while initiation and progression through 25 abortive transcription is slow (Martin and Coleman, 1987; Martin, et al., 1988), elongation of the transcript from 11 to 59 bases is expected to contribute less than 10% to the time required for transcript synthesis. As a consequence 30 the phase of transcription during which TTP is incorporated may not be the rate limiting step in synthesis of the runoff transcript even for the w.t. enzyme. These considerations

imply that one need not expect to see marked differences in synthesis of a relatively short runoff transcript by Y639F or the w.t. enzyme when TTP is substituted for UTP.

Lanes u and v show reactions in which two rNTPs have
5 been substituted with dNTPs (dATP and dCTP), and lanes w and x show reactions in which three rNTPs have been substituted with dNTPs (dATP, dCTP, and TTP). Synthesis of the 59 base runoff transcript is observed with the mutant polymerase (lanes v and x) indicating that Y639F can carry out
10 synthesis even when 3 rNTPs are substituted with dNTPs.

While the abortive transcript patterns in lanes v and x may largely be described as a combination of the patterns observed in lanes h and l there are some important distinctions. The transcript patterns in lanes v and x reveal that the structure of the transcript, as well as the structure of the NTP, affect the rate at which NMPs are added to the transcript. For example, in lanes v and x there is an increase in termination at the 6mer length relative to lane l (the 6mer in lanes v or x runs slightly
15 slower than the poly-G 5mer in the adjacent lane). In lanes l, v and x synthesis of the 7mer involves incorporation of dCMP but in lanes v and x the 6mer contains a dAMP at its 3' end and in lane l it contains an rAMP.

The increase in termination at the 6mer point in lanes v or x indicates that the presence of a dNMP on the 3' end of the transcript reduces the ability of the polymerase to further incorporate NMPs. The high level of termination after synthesis of the rGrGrGdA 4mer in lane h relative to the observed termination after synthesis of the rGrGrGrA
25 4mer in lanes c or d similarly indicates that the presence of deoxynucleotides in the transcript, at least at the 3'-end, influences subsequent extension of the transcript.

Fig. 2 shows the effect of dGTP substitution on transcription by the w.t. and Y639F polymerase. Polymerases, template (supercoiled(Sc) or *HindIII* linearized (Li) pT75), and NTPs were as indicated. Polymerase and template concentrations and electrophoresis conditions as in Fig. 1. Left panel: Labeling was with α -P³² rGTP (a, d, g, j) or α -P³² dGTP(all other lanes). The runoff transcript from the *HindIII*-cut template is indicated in lane f. Right panel: Labeling was with α -P³²-rGTP (a, b, f, g) or α -P³²-dGTP (all other lanes). Poly-rG and poly-dG products of various sizes are indicated in lanes a, c, d. Alignment of these transcript patterns with those in lanes b, c, h, and i in the left panel reveals that the added complexity of the transcript pattern in the latter set of lanes is due to the presence of a mixture of heterogeneous sequence and poly-G transcripts. Heterogenous sequence abortive transcripts are indicated in lane c of the left panel ("4H", "5H").

Fig. 2 reveals the effects of substituting dGTP for rGTP in transcription reactions with the w.t. or Y639F polymerase. In reactions containing only dGTP and *HindIII*-cut pT75 as the template, the mutant polymerase synthesizes poly-dG transcripts up to 4 bases in length (lane c, right panel). Note that--consistent with the assumed structural differences--the poly-dG transcripts ("2dG", "3dG", etc.) in lane c do no co-migrate with the poly-rG transcripts ("2rG", "3rG", etc.) in lane a despite length and sequence identity. When a supercoiled template is used, poly-dG transcripts up to 5 bases in length are obtained (lane d, right panel). In lane e rGMP is added to reactions which contain only dGTP. Ribo-GMP can serve as the initiating, but not elongating nucleotide during transcription(Martin and Coleman, 1989).

With rGMP we therefore ask whether either polymerase can elongate with dGTP if an rNMP is provided for initiation. Addition of rGMP to reactions with dGTP further extends the lengths of the transcripts obtained with the mutant 5 polymerase (lane e, right panel). With the w.t. enzyme very little synthesis is observed in reactions with dGTP (lanes h-j, right panel), though the normal pattern of poly-rG synthesis is observed in the rGTP reactions (lanes f, g, right panel).

10 When reactions contained three rNTPs and dGTP synthesis of runoff transcript from the *HindIII*-cut template is reduced much more than in reactions in which rGTP is present but other ribonucleotides are substituted with deoxynucleotides. For example, there is very little runoff transcript in lane h of the left panel of Fig. 2. Addition 15 of rGMP increases the amount of runoff transcript made by the mutant enzyme in a reaction containing dGTP (lane i) but the amount of runoff transcript is still much less than in reactions with rGTP. On a supercoiled template (lanes a-c, 20 left panel) high levels of long transcripts are obtained with the mutant enzyme in the reactions with dGTP. The w.t. enzyme shows no transcript synthesis in any of the reactions with dGTP irrespective of whether supercoiled templates or rGMP is used..

25 Examination of Fig. 2 shows that the marked reduction in runoff transcript synthesis by the mutant enzyme in reactions with dGTP is not due to a deficit in initiation. In fact, in all of the reactions with dGTP we observe abundant synthesis of 2-~6 base transcripts with the mutant 30 enzyme. The low level of runoff transcript synthesis means that these short transcripts are being inefficiently extended to greater lengths. It should also be noted that

the abortive transcript pattern seen in lanes b, c, h, and i of the left h panel in Fig. 2 is too complex to be accounted for by presuming one product of each base length. Aligning the transcripts produced in the presence of dGTP only with those produced with dGTP, rUTP, rCTP, and rATP allows us to identify the predominant abortive transcripts in lanes c, d, h, and i of the left panel as poly-dG products. In lane c of the left panel we indicate the major non poly-dG abortives by "4H" and "5H". In lanes b, c, h, and i of the left panel of Fig. 2 we therefore see a pattern similar to that of lane h in Fig. 1. The presence of a mixture of normally extended heterogenous sequence and poly-G abortive transcripts indicates that 'normal' transcript extension is inefficient.

It has been shown that mutations in T7 RNAP that reduce phosphodiester bond formation rates cause poly-rG transcript synthesis even when 4 rNTPs are present (Bonner, et al., 1994). It can be seen that the ratio of poly-G to heterogenous sequence transcripts in lanes b, c, h, i of the left panel of Fig. 2 is greater than in lane h of Fig. 1, indicating a greater deficiency in normal transcript extension when dGTP is substituted for rGTP than when dATP is substituted for dGTP. Note that this occurs even though normal extension of the dGdGdG trimer in lane i of the left panel of Fig. 2 (for example) would involve addition of a ribo-AMP while extension of the rGrGrG trimer in lane h of Fig. 1 would involve addition of a deoxy-AMP, clearly highlighting the role of transcript structure, as well as substrate structure, in determining the efficiency of transcript extension. These results show that Y639F can initiate and elongate transcripts with dGTP substituted for rGTP but normal extension of the transcripts in the 2-8 base

range is impaired leading to a large increase in the proportion of poly-dG and short transcripts synthesized. Addition of rGMP to serve as the initiating nucleotide and the use of a supercoiled template both enhance the ability 5 of the mutant to extend the short transcripts during the initial stages of transcription.

Barriers to initiation and extension of the initial transcript with dNTPs: Fig. 2 reveals that Y639F can efficiently transcribe with dGTP the initial G segment of a 10 promoter that initiates "GGGA" but is severely blocked in extending the dG trimer with the A. This implies that the sequence of the initially transcribed region may influence the efficiency with which Y639F can extend the transcript when using dNTPs. Fig. 2 also shows that supercoiling, 15 which presumably facilitates unwinding of the template, enhances the activity of Y639F when using dNTPs. To evaluate the effects of sequence and single-strandedness in the initially transcribed region on the activity of Y639F when using dNTPs, we examined transcription from a set of 20 synthetic promoters which differed in the sequence of their initially transcribed regions and in being fully double-stranded or single-stranded in their initially transcribed regions (Fig. 3).

Fig. 3 shows the effects of single-strandedness and 25 sequence in the initially transcribed region on the activity of Y639F in reactions with 4 rNTPs or 4 dNTPs. Poly-rG transcripts of various sizes are indicated in lane a. Reactions contained the indicated NTPs and polymerases. Polymerase and promoter concentrations were 10^{-6} M and 10^{-5} M, respectively. NTP concentrations and electrophoresis as 30 in Fig. 1. Indicated in some of the lanes are the sequences of different transcripts as deduced from alignment with

(099653A1E-011102)

poly-rG or poly-dG ladders synthesized in the presence of rGTP or dGTP only. The synthetic promoter templates used are double-stranded and have the sequence--

CGAAATTAATACTCACTATA (SEQ ID NO:2)--in their -23 to -1 regions. The promoters differ in their initially transcribed regions as follows: b-e, t, u: GGACT; f-j, v, w: GAGACCGG; a, j-m, x, y: GGGAGACC; n, o, z: GGAAAATT; p-s: GGGGGGGGGGGACT (SEQ ID NO:3). The promoters also differ in being double-stranded (b, d, f, h, j, l, p, r) or single-stranded (other lanes) in their transcribed regions.

For most of the promoters tested, transcription with rNTPs was not markedly affected by having the initially transcribed region be single-stranded. However, when transcribing with 4 dNTPs, Y639F was more active on the partially single-stranded promoters. For example, in lane e (partially single-stranded promoter) of Fig. 3 transcription products are both more abundant and extend to greater lengths than in lane d, where the promoter is fully double-stranded. A similar comparison may be made between lanes m and l or s and r. Regarding sequence we found that a promoter which initiated with 3 G's was superior to a promoter which initiated with two, which was in turn superior to a promoter which initiated with just one G. Thus Y639F activity when using dNTPs was greatest on a promoter which initiates "GGGAGACC" (lanes l and m). The initially transcribed region of this promoter corresponds to the consensus sequence for T7 promoters in the +1 to +6 segment. This was the only promoter which, when fully double stranded, gave rise to high levels of transcript synthesis with Y639F in reactions containing only dNTPs (lane l).

Promoters which initiated with 2 G's (lanes d, e, o) gave lower levels of transcript synthesis in the 4 dNTP reactions, and a promoter which initiated with 1 G (lanes h, i) was not utilized by Y639F in reactions with 4 dNTPs.

5 Within the initially transcribed region, elements other than the number of G's appear to be important. For example, we have found that the w.t. or Y639F polymerases are less efficient in initial transcription extension on the T7 promoter found in the pBS plasmid which initiates GGGC, than

10 on the ϕ 10 promoter found in pT75 which initiates with the consensus GGGAGA (data not shown).

Another example is evident in lanes n and o which show the transcripts obtained on a partially single-stranded promoter that initiates GGAAAAUU. Like the promoter used in lanes c and e, this promoter initiates with 2 G's but normal transcript extension on this promoter is less efficient than on the promoter that initiates GGACU. In the 4 rNTP reactions (lanes c, n), the proportion of short transcripts (dimers, trimers) is greater in lane n and we observe significant amounts of poly-rG transcripts beyond the dimer length in lane n but not in lane c. In the 4 dNTP reactions almost all of the transcripts in lane o terminate at the trimer or tetramer length. Because increasing the number of Gs from 1 to 3 enhanced Y639F activity when using dNTPs, we

15 tested a promoter which initiated with a run of 11 G's (lanes p-s). A potential drawback to such a promoter is that such a long run of G's could inhibit the ability of the polymerase to unwind the template. Since T7 promoters with more than 3 consecutive G's in the initial region do not

20 occur naturally, it may be that for other reasons such sequences do not favor initial transcript extension. In fact, we find that this promoter is a poor template. When

25

30

it is fully double-stranded, initiation and extension of the transcript is inefficient with either rNTPs (lane p) or dNTPs (lane r), consistent with the expectation that a promoter with this sequence would be difficult to melt.

5 Initiation and transcript extension is enhanced when this
promoter is partially single-stranded (lanes q and s), but
while poly-G transcripts from 2 to 7 or more bases in length
are abundant, runoff transcripts of the expected length are
not predominant products. In reactions with 4 rNTPs the
transcript patterns of the w.t. and Y639F polymerases are
virtually identical so we do not repeat the 4 rNTP reactions
with the w.t. enzyme in Fig. 3. Lanes t-z show that the
w.t. enzyme is virtually inactive in reactions with 4 dNTPs
and the same set of promoter used in lanes b-s.

15 Relative selectivity of the mutant and w.t. polymerases
for dNTPs and rNTPs: Figs. 1-3 present a qualitative
analysis of the structure of the transcripts produced by the
w.t. and Y639F polymerases with various combinations of
NTPs. They show that Y639F can use dNTPs with high
efficiency and that both transcript and substrate structure
play a role in determining the efficiency of transcript
extension. To obtain a quantitative measure of the relative
selectivity of the w.t. and mutant polymerases for dNTPs vs.
rNTPs under conditions where transcript structure was not a
complicating factor, we carried out reactions in which all 4
rNTPs were present but rNTPs or dNTPs were used to
radiolabel the transcripts (Table I, see Appendix 1). Under
these conditions the unlabeled rNTPs were present in vast
excess relative to the labeling NTPs so labeling dNTPs are
almost always incorporated adjacent to rNTPs and into
transcripts of nearly uniform rNMP structure. On average
the mutant enzyme is ~20-fold less selective for rNTPs over

DRAFTED BY [REDACTED]

dNTPs than the w.t. enzyme. We used this assay to screen our collection of T7 RNAP active site mutants for increased dNTP utilization. In this screen we looked for increased dATP incorporation in transcription reactions with all of 5 these mutants using supercoiled pT75 as the template, 4 cold rNTPs, and P³²-rATP or P³²-dATP to label. Since these results were negative with the exception of the tyrosine 639 10 mutants, we do not present them here, but the mutants tested in this way are listed in "Materials and Methods". It has been reported that use of Mn⁺⁺ instead of Mg⁺⁺ decreases 15 substrate discrimination and increases miscoding for a number of polymerases (Tabor and Richardson, 1989; Nivogi and Feldman, 1981). We, therefore, examined the rNTP/dNTP selectivity of the mutant and w.t. enzymes in Mn⁺⁺-citrate buffer (Tabor and Richardson, 1989). With Mn⁺⁺ the preference of both the w.t. and Y639F polymerases for rNTPs over dNTPs was markedly reduced.

Relative activity of the w.t. and Y639F polymerases with different NTP combinations: The relative activities of 20 the Y639F and w.t. polymerases with supercoiled pT75 as a template and various combinations of rNTPs/dNTPs were measured in both Mg⁺⁺ and Mn⁺⁺ buffers (Table II, see Appendix 1). In Mg⁺⁺ buffer, substitution of a single rNTP with a dNTP reduces w.t. activity by 20 to >400-fold, but 25 only modestly reduces the activity of Y639F. The rank order of the effect of a particular dNTP substitution on w.t. enzyme activity--dGTP>dATP>dCTP>dTTP=dUTP--matches the order of their addition to the transcript. With the "2 dNTP" reactions the w.t. enzyme was most active when rCTP and rUTP 30 were substituted with dNTPs, corresponding to the 2 nucleotides added latest to the transcript. The w.t. enzyme was inactive with all other "2 dNTP" or "3 dNTP". In the "2

dNTP" reactions Y639F was least active in the dGTP, dATP reaction, corresponding to the 2 nucleotides incorporated first during transcription. In the "3 dNTP" reactions Y639F was least active in the dGTP, dATP, dCTP reaction,
5 corresponding to the 3 nucleotides incorporated first during transcription.

In Mn⁺⁺ buffer both the w.t. enzyme and Y639F show a reduction in their sensitivity to substitution of dNTPs for rNTPs, consistent with an expectation of reduced substrate 10 discrimination in Mn⁺⁺ buffer. There was, however, also a sharp reduction in overall activity with Mn⁺⁺. We varied Mn⁺⁺ concentrations over a wide range (from 20 mM to 150 μM in 2-fold dilutions) to determine if an optimal Mn⁺⁺ concentration that would result in high activity could be 15 identified, but we found similar activity at all Mn⁺⁺ concentrations tested (data not shown). Thus, while discrimination between rNTPs and dNTPs was less in Mn⁺⁺ buffer, Y639F is more active in Mg⁺⁺ buffer than in Mn⁺⁺ buffer with all NTP combinations examined. Similarly, the 20 w.t. enzyme exhibits greatly reduced discrimination between rNTPs and dNTPs in Mn⁺⁺ buffer, but is modestly more active in Mn⁺⁺ buffer than in Mg⁺⁺ buffer only for certain combinations of dNTPs and rNTPs.

DNA and RNA synthesis on homopolymeric templates: T7
25 RNAP will synthesize poly(rG) RNAs on poly(dC) templates (Bonner, et al., 1994; Ikeda and Richardson 1987). We measured the activity of the w.t. and mutant polymerases on poly(dI)•poly(dC) with rGTP, dGTP, dGTP+rGMP (Table III, see Appendix 1). The activity of T7 RNAP on poly(dI)•poly(dC)
30 and poly(dC) is especially robust. Mutant polymerases that have greatly reduced activity on normal promoter templates still display high activity on poly(dC) templates (Bonner,

et al., 1994). We, therefore, characterized two poorly active non-conservative tyrosine mutations on this template (Y639A and Y639S). In Table III we also present results obtained with mutant G640A. Presentation of data for the 5 latter mutant was selected because it is more comparable in activity to the Y639A/S mutants and because it is representative of a mutation which has marked effects on the kinetics of transcription but does not affect substrate discrimination, even though it is directly adjacent to Y639. 10 We find that all of the Y639 mutants exhibit reduced substrate discrimination as demonstrated by the fact that their differential activity in reactions containing dGTP or dGTP+rGMP vs. rGTP is less than for the w.t. enzyme or the G640A mutant. In fact Y639F displays similar activity with 15 rGTP, dGTP, or dGTP+rGMP.

Since the nucleic acid synthesized by Y639F on poly(dI)•poly(dC) using dGTP is presumably composed solely of dNMPs it is expected to be resistant to alkaline hydrolysis (Schmidt and Tannhauser, 1945). Fig. 4 shows 20 transcription by Y639F and w.t. polymerase with dGTP or rGTP on poly(dI)•poly dC). Transcription reactions were carried out with poly(dI)•poly(dC) at .2 mg/ml and the indicated NTPs and polymerases. Reaction products were left untreated (-) or treated with 1 M NaOH for 5 hours at 37°C (+). 25 Polymerase and NTP concentrations and electrophoresis as in Fig. 1. Labeling was with α -P³² rGTP (a, d, g, j) or α -P³² dGTP (other lanes). Fig. 4 shows the transcription products obtained with the w.t. or Y639F polymerases on poly(dI)•poly(dC) before and after treatment with alkali. 30 With dGTP or dGTP+rGMP the w.t. enzyme is poorly active in the synthesis of long transcripts, however we can observe smears of heterogeneously sized, short transcripts in the

09955376 - 047202

reactions with the w.t. enzyme and the dGTP substrates (lanes e and f) while Y639F synthesizes higher levels of long transcripts which are retained near the top of these gels (lanes b and c). The presence of these short 5 transcripts indicates that Y639F and the w.t. enzyme differ only in the degree to which they can utilize dNTPs. The w.t. can also initiate and extend transcripts with dGTP, but it is much less processive when using dNTPs than Y639F so its transcripts are much shorter. When these reactions are 10 treated with base, degradation of the long transcripts made in the reactions with rGTP is observed and the amounts of short RNAs (presumably hydrolysis products) increase (lanes g and j). In the reactions in which dGTP or dGTP+rGMP were used as substrates no degradation of the transcripts by base 15 treatment is observed, confirming that these transcripts are composed of dNMPs.

T7 RNAP as a reverse transcriptase or RNA replicase which initiates *de novo*: It has been reported that T7 RNAP can use both RNA and DNA templates (Konarska and Sharp, 20 1989). We, therefore, determined if the Y639F mutant would use dNTPs when transcribing an RNA template (poly(rC), Table IV). Overall the activity of the w.t. and Y639F polymerases on poly(rC) with rGTP was 10-20-fold less than on poly(dC) (not shown), but this reduction did not preclude synthesis 25 of high levels of RNA on poly(rC) by using higher polymerase concentrations than were used in the poly(dI)•poly(dC) reactions. When dGTP or dGTP+rGMP was used the w.t. enzyme was not measurably active on poly(rC), while the activity of Y639F was reduced by only ~4-fold (with dGTP+rGMP) or ~8-fold (with dGTP). Thus, both the w.t. and Y639F polymerases 30 are capable of unprimed RNA-directed RNA polymerization

while Y639F is also capable of unprimed reverse transcription.

DNA- and RNA-primed synthesis of DNA and RNA: In the assays described so far we have examined *de novo* initiated synthesis. T7 RNAP can also extend RNA primers. We, therefore, examined the abilities of both polymerase to carry out DNA or RNA primed synthesis of DNA and RNA (Fig. 5).

Fig. 5 shows primed synthesis of DNA and RNA with Y639F and the w.t. polymerase. Transcription reactions contained end-labeled 12 base DNA or RNA primers of identical sequence (GGACACGGCGAA, SEQ ID NO: 4) hybridized to a DNA template (CCCGGGATGGAATGGAGTATTGCCGTGTCCATGGCTGTAAGTATCC, SEQ ID NO: 5). Primer-template concentration was 10^{-5} M. Reactions contained the indicated polymerases (10^{-7} M) and NTPs. NTP concentrations and electrophoresis as in Fig. 1.

We found that both the w.t. and Y639F polymerases can extend DNA and RNA primers with rNTPs, but extension of DNA primers was 2-3-fold less efficient than extension of RNA primers. Y639F also extended DNA and RNA primers with dNTPs but ~4-fold less efficiently than with rNTPs.

The Y639F mutant does not exhibit greatly increased miscoding: We examined the miscoding properties of the w.t. and mutant T7 RNAPs by measuring the relative incorporation of labeled rGTP, rUTP, rATP, and rCTP on poly(dC) or poly(dT) templates in the presence of excess unlabeled rGTP or rATP, respectively (Table V, see Appendix 1). An increase in miscoding would be reflected in an increase in the rate of incorporation of the non-complementary NTP into RNA.

On poly(dC), the w.t., Y639F, and G640A polymerases incorporate rGTP into RNA at greater than 1300-2000-fold the

00000000000000000000000000000000

rate of rUTP incorporation. Ribo-GTP is incorporated some
400-600-fold better than rCTP, and 200-400-fold better than
rATP. Because of their lower activity we can say only that
the relative rate of incorporation of rGTP on poly(dC) is
5 184-fold greater than the rate of incorporation of non-
complementary rNTPs for Y639A, and 50-fold greater for
Y639S. The use of Mn⁺⁺ instead of Mg⁺⁺ has been reported to
increase miscoding for a number of polymerases (Tabor and
Richardson, 1989; Nivogi and Feldman, 1981) so we examined
10 the effects of Mn⁺⁺ on miscoding by w.t. and Y639F
polymerases. In Mn⁺⁺ buffer the G640A, Y639A, and Y639S
mutants were insufficiently active to allow accurate
measures of miscoding. With the w.t. and Y639F polymerases
the use of Mn⁺⁺ increases miscoding by 20-40-fold. However,
15 the apparent rate of miscoding by Y639F remains similar to
the w.t enzyme.

On poly(dT), high levels of activity allowing an
accurate measure of miscoding frequencies could only be
observed for Y639F and the w.t. polymerase in both Mg⁺⁺ and
20 Mn⁺⁺ buffers. In Mg⁺⁺ buffer apparent miscoding rates on
poly(dT) were higher than on poly(dC), but were similar for
Y639F and the w.t. enzyme. In Mn⁺⁺ buffer miscoding rates
were increased by ~5-fold, on average, but again rates were
similar for w.t. and Y639F. However, on poly(dT) Y639F did
25 show a reproducible ~2-fold increase, relative to the w.t.
enzyme, in the ratio of the rates of rGTP to rATP
incorporation.

Homopolymer assays have been used previously to measure
miscoding by RNAPs (Nivogi and Feldman, 1981; Glazer, 1978;
30 Blank, et al., 1986) but it should be remarked that they can
produce only upper bounds for miscoding frequencies.
Measured miscoding rates could reflect contamination of the

homopolymeric templates. It is also possible that the transcripts themselves could serve as templates and support incorporation of rNMPs non-complementary to the original template (i.e., the poly(rG) or poly(rA) transcripts made on poly(dC) or poly(dT) could subsequently support synthesis of poly(rC) or poly(rU) transcripts). Such caveats are less relevant to the miscoding observed in Mn⁺⁺ buffer since the change in divalent cation increases miscoding but cannot affect template composition. However, in Mg⁺⁺ buffer we should consider the measured miscoding frequencies to be upper bounds for the true rates of miscoding. Nevertheless, the results presented in table V indicate that Y639F does not exhibit a gross increase in miscoding which would manifest itself as a clear increase in the incorporation of non-complementary rNMPs on homopolymeric templates.

The increased utilization of dNTPs by Y639F is due to both a decreased K_m and an increased k_{cat} for dNTPs: The K_m and k_{cat} of the w.t. and Y639F polymerases with rATP, rITP and 5 different dNTPs were measured (Table VI, see Appendix 1). The K_m of the w.t. enzyme for dNTPs was much higher than previously reported values for the corresponding rNTPs and varied considerably for different dNTPs (Ikeda and Richardson, 1987; Patra, *et al.*, 1992). Notably the w.t. enzyme K_m values correlate with the rNTP/dNTP selectivity values presented in Table I. The selectivity of the w.t. enzyme with ribo- vs. deoxy-nucleotides was greatest for CTP, followed by ATP, and was the least for UTP, implying that an important component of the selectivity of the w.t. enzyme for rNTPs over dNTPs is a much higher K_m for dNTPs. For Y639F, the K_m values for these dNTPs are from ~3 to ~11-fold less, but the rank order of these K_m values (dCTP K_m > dATP K_m > dGTP K_m > dTTP K_m) is the same as for the w.t.

enzyme. For Y639F, k_{cat} values for reactions with different dNTPs were only 2-4 fold less than for rNTPs, while the w.t. enzyme displayed k_{cat} values with dNTPs that were from ~6 to ~30-fold less than for rNTPs.

5 Elongation rates of Y639F in reactions containing a single dNTP: Elongation rates for Y639F in reactions with 4 rNTPs or 1 dNTP and 3 rNTPs were determined by analyzing aliquots, taken at 10 second intervals, from transcription reactions initiated on supercoiled pT75 by adding NTPs to 10 otherwise complete reaction mixes (Fig. 6).

Fig. 6 shows relative elongation rates of Y639F in "4 rNTP" and "3 rNTP + 1 dNTP" reactions. The template was supercoiled pT75 at 5×10^{-7} M. Y639F polymerase was used at a concentration of 10^{-6} M. Reactions contained the indicated NTPs. Labeling was with α -P³²-rATP (lane e) or α -P³²-rGTP (other lanes). After initiation of the reactions aliquots were taken at 10 second intervals and analyzed on 1% agarose denaturing-formaldehyde gels. The figure shows the 20 second time point. The bars indicate the positions 20 of λ DNA markers.

When analyzed on denaturing agarose gels (Fig. 6) the heterogeneously sized transcripts from these reactions are resolved as a smear with the trailing edge of the smear corresponding to transcripts initiated at t=0 and from which the maximal transcript elongation rate can be determined (Golomb and Chamberlin, 1974; Bonner, *et al.*, 1994). The following elongation rate reductions (relative to a '4 rNTP' reaction) were obtained for reactions containing a 3 rNTPs and 1 dNTP:dATP, ~3-fold; dUTP, ~2-fold; dGTP ~1.5-fold; 25 dCTP, 1-1.5-fold. Because of its poor activity in reactions with dNTPs we could not determine the corresponding elongation rate reductions for the w.t. enzyme.

Other non-canonical nucleoside triphosphate substrates
for mutant polymerases: Although wild-type T7 RNAP can not
efficiently utilize dideoxy-NTPs as substrates, we have
found that Y639 mutants of this enzyme can also use dideoxy-
5 NTPs as substrates (Table VII). We have also found that
Y639 mutants can use other non-canonical nucleoside
triphosphates as substrates (Table VIII). Nucleic acids
synthesized by incorporation of some non-canonical
nucleotides, such as 2'-F-NTPs, may offer advantages in
10 being more resistant to digestion by nucleases such as
ribonucleases. Other uses and advantages of various non-
canonical nucleotide substrates in methods of the present
invention using the mutant polymerases will be apparent
after examination of the specification, claims and drawings.

15 Other mutations: It is conceivable that, in the
absence of a bound rNTP, a hydrogen bond forms between Y639
and some other active site side chain. The possibility of
an interaction between M635 and Y639 was tested since M635
and Y639 are close and M635 approaches the ribose in our
20 models of NTP in T7 RNAP (Huang, *et al.*, submitted for
publication). This position is methionine in the T7 RNAP
class of RNAPs (McAllister, W.T., 1993), but is either
tyrosine or phenylalanine in the homologous DNAPs, and in
the DNAP mutants at this site (i.e., positions homologous to
25 position 762 in *E. coli* DNAP I) that affect dNTP/ddNTP
discrimination (Tabor, S., and Richardson, C.C., European
patent application, 1994). While M635A, M635F or M635Y
mutants had effects on NTP K_m, they did not affect 2'-group
discrimination with respect to dNTPs and rNTPs in either
30 wild-type or the Y639F T7 RNAP background. Additional
studies will reveal whether these M635 mutations have other
effects, such as effects on discrimination at the 3'-

position of the sugar, or effects on discrimination with respect to NTPs with other substituents, such as fluorine at the 2' position of the sugar. Also, studies on similar double mutations at the homologous sites in the homologous 5 class I DNAPs will reveal the effects of such double mutations on discrimination at the 2'- and 3'-positions of the sugar; specifically, these studies will reveal whether class I DNAPs having a phenylalanine at the position homologous to amino acid position 766 in *E. coli* DNAP I have 10 reduced discrimination for rNTPs if the amino acid is methionine or tyrosine at the position homologous to amino acid position 762 in *E. coli* DNAP I.

Provided that class I DNAP mutants which have a reduced discrimination for rNTPs compared to dNTPs can be obtained, 15 it will be possible to use such DNAP mutants to carry out the methods of the present invention that comprise nucleic acid synthesis from a nucleic acid primer, at least part of which is sufficiently complementary to a template nucleic acid to hybridize therewith and to be extended by the 20 polymerase. An especially preferable use for such mutant DNA polymerases would be to carry out Partial Ribo- substitution sequencing reactions from primers, whether labelled by any of the methods known in the art, or unlabelled. Since the sequence-delimiting rNTP nucleotides 25 for the Partial Ribo-substitution Reaction do not terminate the growing phosphodiester chain when they are incorporated during nucleic acid synthesis, the DNA synthesis for Partial Ribo-substitution can occur simultaneous with and be identical to nucleic acid synthesis for another procedure, 30 such as NASBA, 3SR, TMA or another similar method, provided that the mutant DNAP with reduced discrimination for rNTP compared to dNTPs is thermostable, or the PCR strand

[REDACTED]

displacement amplification or other methods for nucleic acid amplification involving nucleic acid synthesis from a primer. The nucleic acid products containing the sequence-delimiting rNMPs can then be cleaved by treatment with a 5 chemical base or a ribonuclease, and analyzed by methods known in the art to obtain the nucleotide-specific pattern of bands or, provided that a Partial Ribo-substitution Reaction is carried out for each of the four nucleotides, the complete sequence of the nucleic acid. Also, provided 10 that primers with distinguishable non-radioactive labels are used, multiple Partial Ribo-substitution sequencing reactions may be carried out simultaneously in the same reaction mixture and run and read in the same lane of a polyacrylamide gel or capillary tube or other matrix for 15 separating the fragments based on size, as is the case for all of the methods of the present invention described herein.

C. Discussion

Our results reveal that mutations of tyrosine 639 in T7 20 RNAP reduce the ability of the polymerase to discriminate between rNTPs and dNTPs. A conservative mutation which removes the tyrosine hydroxyl but retains the phenolic ring (Y639F) exhibits w.t. activity but an average reduction of ~20-fold in the selectivity for dNTPs over rNTPs (Table I). 25 Non-conservative mutations of this tyrosine (Y639A/S) also display decreased rNTP/dNTP discrimination (Table III), but are less active than the w.t. enzyme. Replacement of an rNTP by a dNTP typically reduces Y639F transcript elongation rates by only a factor of two. Tyrosine 639 is conserved in 30 a large number of DNA-directed RNA and DNA polymerases (Delarue, *et al.*, 1990). In DNAP I, mutations of the Y766 to serine and phenylalanine have been characterized

(Polesky, et al., 1989; Carroll, et al., 1991). The Y766S mutation was alone amongst a number of active site mutations characterized in decreasing DNAP I fidelity (increasing miscoding). The Y766F mutation displayed w.t. fidelity and activity. Similarly, the T7 RNAP Y639F mutant displays w.t. kinetics (Bonner, et al., 1992, 1994; Woody, et al., 1994), and the only effect we can identify for this mutation in T7 RNAP is the reduced substrate discrimination reported here. Thus, while T7 RNAP Y639F showed decreased dNTP/rNTP selectivity, it did not exhibit increased miscoding as assessed by incorporation of non-complementary NTPs on homopolymeric templates.

The Y639 T7 RNAP mutations present us with, in one sense, the functional unification of polymerases to go along with their structural unification. The active site of w.t. T7 RNAP is forgiving with regards to template structure (RNA or DNA) or mode of initiation (primed or *de novo*). A mutation which relaxes the substrate selectivity of this polymerase further expands the range of activities which it can display *in vitro*. Depending on the substrates and templates presented to it, the Y639F T7 RNAP can act as an RNA- or DNA-directed RNA or DNA polymerase in primed or *de novo* initiated reactions. Thus it can display a variety of activities normally associated with distinct polymerases, including some entirely novel activities such as *de novo* initiated reverse transcription or mixed dNMP/rNMP polymer synthesis.

Example 2: A mutant SP6 RNA Polymerase as a DNA Polymerase

After the observations made above with T7 RNAP, we decided to examine bacteriophage SP6 RNA polymerase to determine whether the DNA synthesis properties observed for the mutant T7 RNAP could, as expected, be extended to other

mutant polymerases. Bacteriophage SP6 is a lytic phage which infects the bacterial species *Salmonella typhimurium* (Butler and Chamberlin, 1982). SP6 phage resembles *E. coli* phage T7 and their genomes are comparable in size, gene organization and pattern of gene expression (Kassavetis, *et al.*, 1982).

The phage encoded RNA polymerases are very similar in size (Butler and Chamberlin, 1982) and amino acid sequence (Katani, *et al.*, 1987).

10 The homologous tyrosine at position 639 in T7 RNA polymerase is readily identified at position 631 in SP6 RNA polymerase (Fig. 7). Substitution of tyrosine 631 with phenylalanine in the SP6 RNA polymerase was expected to confer the same phenotypic changes in catalytic properties
15 in this enzyme as were demonstrated for Y639FT7 RNA polymerase (Example 1).

20 Localized Mutagenesis. Refer to Fig. 7 for a summary of the amino acid and nucleotide sequence surrounding TYR631 in the SP6 RNA polymerase gene. Mutagenesis of the Y631 residue may be accomplished by the method of Kunkel, *et al.* (1991). Alternatively, one of the many other methods for mutagenesis known to those of skill in the art may be used. The amino acid and nucleotide sequences of the resulting TYR631PHE mutant SP6 RNA polymerase are also given in Fig.
25 7. As shown, the A residue at position 2 in codon 631 of the SP6 RNA polymerase gene was changed to a T. This results in the loss of the single *NdeI* restriction enzyme site which is present in the wild-type gene, permitting identification of mutant clones.

30 Preparation and Purification of Mutant RNA Polymerase

A single clone was selected in which the *NdeI* site was missing and which expressed SP6 RNA polymerase activity.

Several liters of LB + Amp were grown from a pUC18 Y631F SP6 RNA polymerase clone overnight at 37°C. Cells were harvested, lysed and Y631F mutant SP6 RNA polymerase was purified approximately according to standard methods (Butler and Chamberlin, 1983).

5 Transcription Reactions

To verify that Y631F SP6 RNAP had the desired phenotype, *in vitro* transcription reactions were done where one of the four rNTPs was substituted by the corresponding dNTP (Sousa and Padilla, 1995). As expected, the Y631F SP6 RNAP mutant displayed reduced dNTP/rNTP discrimination compared with wild-type SP6 RNAP, similar to that observed for the Y639 mutant of T7 RNAP.

In a standard *in vitro* transcription reaction using the 15 four ribonucleoside triphosphates (rATP, rGTP, rCTP, and rUTP), both enzymes, the wild-type SP6 RNA polymerase as well as the Y631F mutant, synthesized the correct 1.4 kb transcript and in the expected amounts as visualized on gels. However, if one of the four ribonucleoside 20 triphosphates, rGTP for example, is completely substituted by dGTP and *in vitro* transcription reactions are done with the wild-type and mutant enzymes, no transcript is made by the wild-type enzymes. However, the mutant enzyme makes the expected full length transcript in good yield as observed on 25 agarose gels..